BIONUMERICS Tutorial:

Importing and processing VNTR sequencer curve files

1 Aim

Comprehensive tools for processing electrophoresis fingerprints, both from slab gels and capillary sequencers are incorporated into BIONUMERICS. When fingerprints are run on a capillary sequencers the resulting data can have two different formats: curve files (also referred to as electropherograms, chromatogram files or trace files), or peak tables. In this tutorial we will focus on the import of curve files. Since curve files contain the raw data, the fingerprint processing steps (e.g. normalization, band assignment, ...) are also covered in this tutorial.

2 Sample data

Raw curve files from Applied Biosystems and Beckman (now AB SCIEX) sequencers can be imported directly in BIONUMERICS. A batch of Applied Biosystems curve files can be downloaded from the Applied Maths website (`https://www.applied-maths.com/download/sample-data`, click on "VNTR sequencer trace files"). These example files will be used to illustrate the import steps. Eight VNTR targets per sample were amplified: two pools were generated (**MP1** and **MP2**), each containing four PCR products, using four color dyes (**6-FAM**, **VIC**, **NED**, and **PET**).

3 Importing curve files

1. Create a new database (see tutorial "Creating a new database") or open an existing database.

2. Select ***File*** > ***Import...*** (, **Ctrl+I**) to call the *Import* dialog box, choose ***Import curves*** under ***Fingerprint type data*** and press <***Import***>.

3. Browse to the downloaded and unzipped example data folder `VNTR sequencer trace files` and select all Applied Biosystems curves files (extension `.fsa`). Press <***Open***>.

The files are displayed in the *Input* wizard page and the default suggested ***Fingerprint file name*** is the folder name.

4. Change the name to **Batch1** and press <***Next***> (see Figure 1).

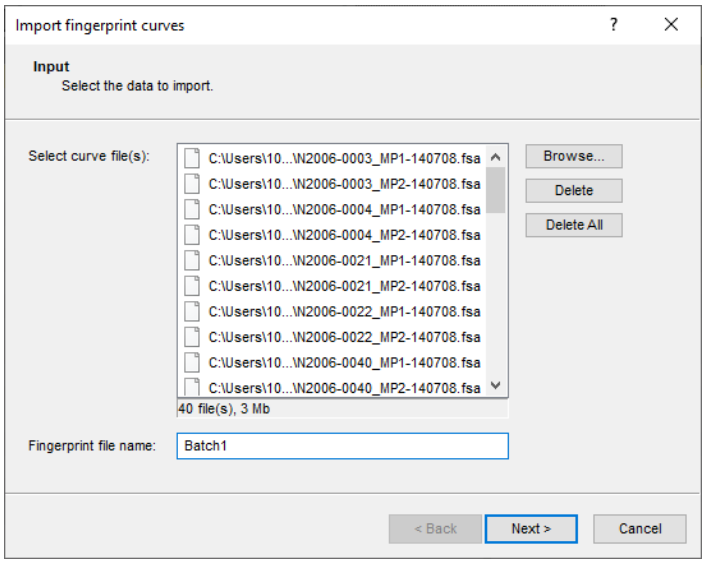

Figure 1: Selected curve files and fingerprint file name.

The way the information should be imported in the database should be specified with an import template. In the example data set, the dye name can be parsed from the file information and the sample and pool information can be parsed from the file names.

5. Select the predefined template ***Example import with pools*** and press the <***Preview***> button to check the rules applied on the selected files.

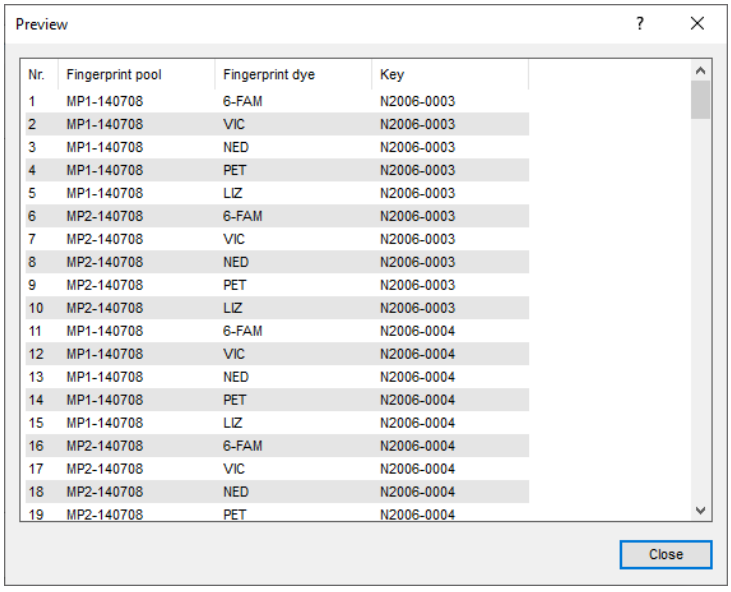

Figure 2: Preview of predefined template.

In the preview we see that the part of the file name appearing before the underscore is linked to the ***Key*** field. The curve dyes are also correctly linked to the ***Fingerprint dye*** type. In a next step, we will modify the rule for the ***Fingerprint pool***: we will link the part of the file name appearing after the underscore and before the hyphen (”-”) to the ***Pool*** field.

6. Close the preview.

7. Make sure the predefined template ***Example import with pools*** is still selected and press the <***Edit***> button to edit the rules.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

8. Check the option ***Show advanced options***, make sure the second row is selected in the grid panel and press the <***Edit parsing***> button.

9. In the *Data parsing* dialog box, update the data parsing string: "*_[DATA]-*". The asterisk will serve as wildcard.

10. Press the <***Preview***> button and press <***OK***> when the parsing is correct (see Figure 3).

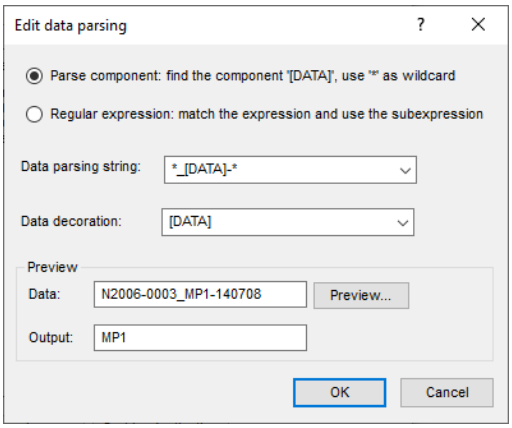

Figure 3: Parsing string.

The *Import rules* dialog box is updated.

11. Press the <***Preview***> button to check the outcome of all defined import rules (see Figure 4).

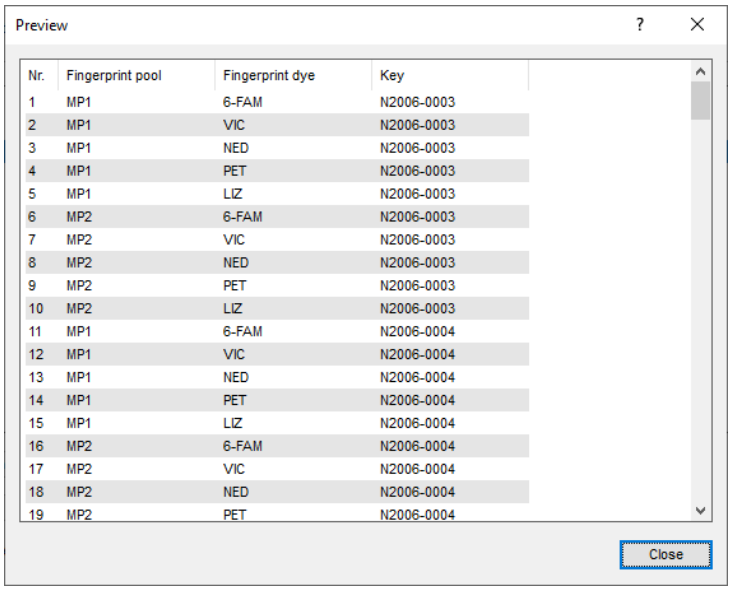

Figure 4: Preview of the import rules.

12. Close the preview and press <***Next***> to go to the next step.

In the example data set, the LIZ channel contains the size standard (GeneScan 500 LIZ).

13. Make sure ***LIZ*** is selected as ***Reference dye*** and press <***Next***> and <***Finish***>.

The predefine import template is now updated.

14. Make sure <***Create new***> is selected as fingerprint type experiment, make sure the ***Example import with pools*** template is selected and press <***Next***>.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

15. Specify a name for the new base fingerprint type experiment (e.g. **MLVA**) and press <***OK***> and confirm.

16. Confirm the proposed OD range.

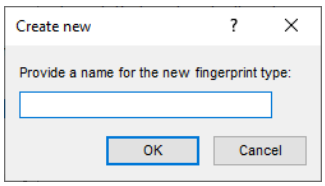

Figure 5: New base fingerprint type experiment.

A fingerprint type needs to be present in the database for each pool and dye combination. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. If one or more of these fingerprint types are not present in the database, a new dialog box pops up, listing the missing fingerprint types (see Figure 6).

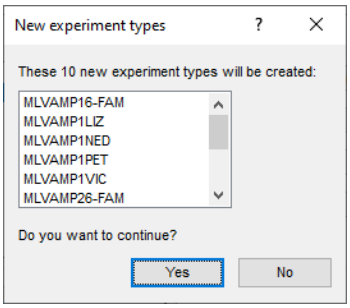

Figure 6: Missing fingerprint experiments.

17. Confirm the creation of the missing fingerprint type experiments.

18. Press <***Next***> to confirm the creation of new entries in the database.

19. Make sure ***Open curve preprocessing window*** is checked in the last step and press <***Finish***>.

For each dye checked in the *Dyes panel* of the *Import data* dialog box, a new fingerprint file is created, composed of the file name specified and the name of the dye (e.g. Batch1_LIZ). These files are displayed in the *Fingerprint files* panel. The reference file is shown in the **Link** column. Double-clicking on a fingerprint file opens the *Fingerprint* window. If lane information was imported with the individual lanes, this information is displayed in the *Fingerprint information panel*.

The imported fingerprint lanes are linked to new entries in the database. The lanes are linked to the corresponding fingerprint ”dye” type. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. The fingerprint type experiments are displayed in the *Experiment types* panel.

Entries for which fingerprint data was imported are selected in the database.

After data import, the *Main* window looks as in Figure 7.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

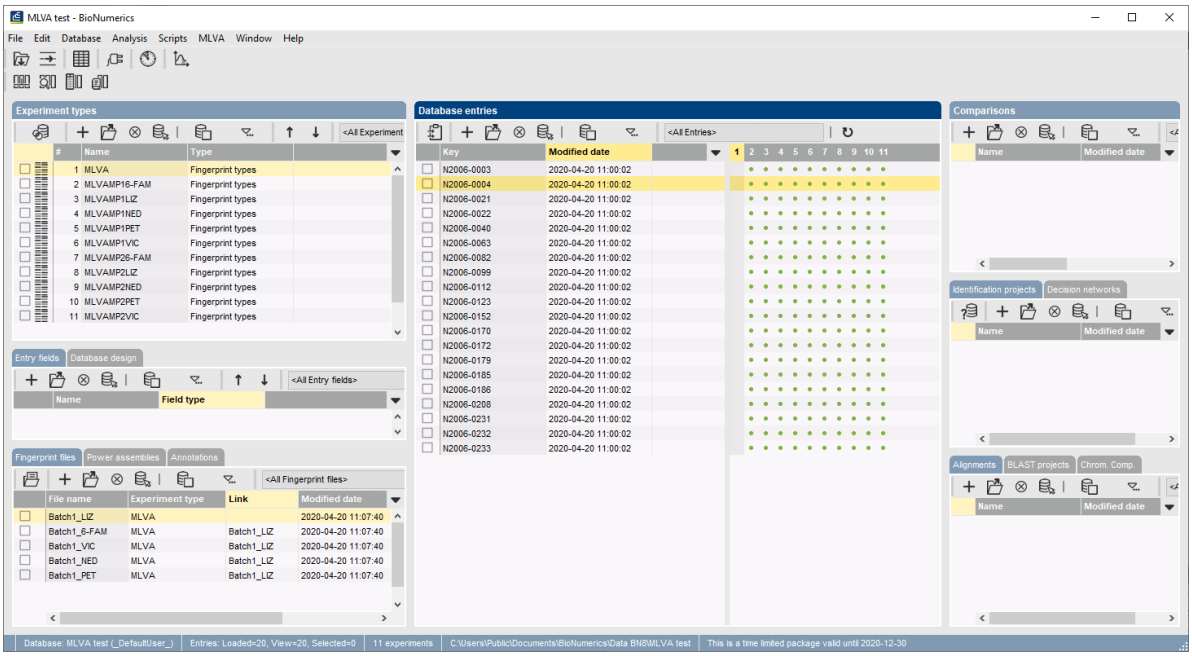

Figure 7: The *Main* window after import of the data.

4 Processing curve files

When the option ***Open curve preprocessing window*** was checked in the last step of the import routine, the *Fingerprint curve processing* window opens when pressing the <***Finish***> button. All channels from the run are automatically loaded and displayed in the *Fingerprint curve processing* window.

The *Fingerprint curve processing* window can also be called from the *Main* window by highlighting one of the channels in the *Fingerprint files* panel and selecting ***Open fingerprint data...*** (). Alternatively, you can first open the *Fingerprint* window with ***Edit*** > ***Open highlighted object...*** (, **Enter**) and then select ***File*** > ***Edit fingerprint data...*** ().

1. Click on the icon left of the data channels in the *Channels* panel.

The data channels are now hidden from the view and its icons are displayed as .

2. Use the zoom sliders on the left and on top to optimize the display of the fingerprint curves.

Since the raw chromatogram files have not undergone any preprocessing, normalization will have to be performed. This requires a *reference system* to be defined, based upon the marker peaks available in the reference dye.

3. Make sure the reference dye is the only dye visible in the upper panel.

4. Select ***Bands*** > ***Search reference bands...*** (, **Ctrl+F**) to call the *Search reference bands* dialog box.

5. Leave the default settings unaltered and press <***OK***>.

The bands that fall within the specified criteria are marked with a solid line at the band's position (see Figure 8).

6. To have a reference system automatically created based on a lane containing commercial size marker, first highlight a suitable lane and then select ***References*** > ***Define size standard...***.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

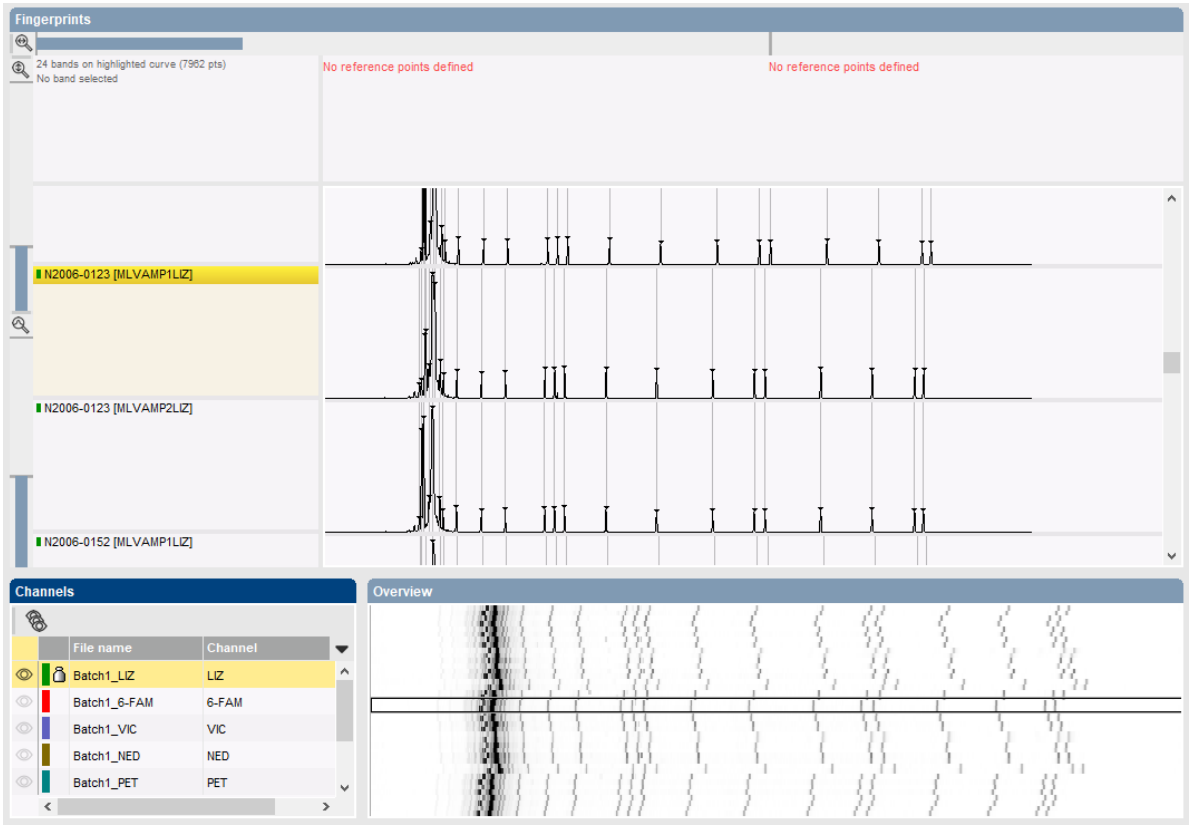

Figure 8: The *Fingerprint curve processing* window only displaying the reference dye.

This will display the *Size standard* dialog box, from which a size marker can be selected (see Figure 9).

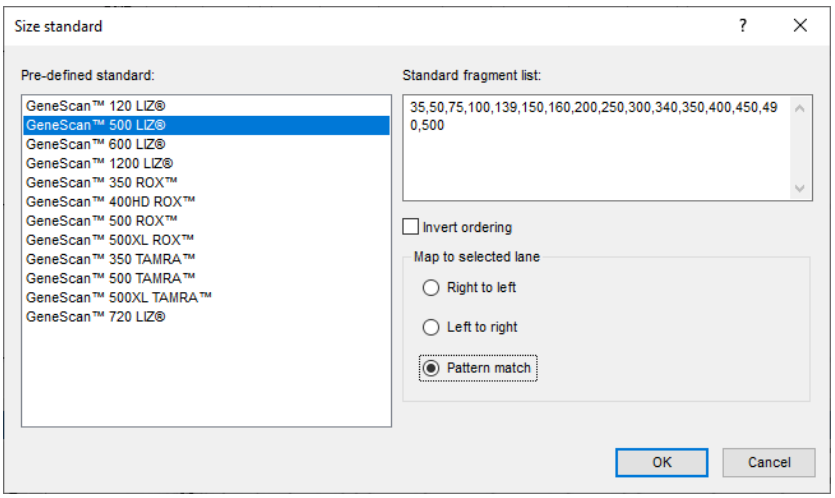

Figure 9: Select marker from the list.

In the example curve files, the GeneScan 500 LIZ size standard is used as reference, containing 16 bands with known molecular weight.

7. Select **GeneScan 500 LIZ** from the list, select **Pattern match** and press <***OK***> twice.

8. Save the data to the database with ***File*** > ***Save*** (, **Ctrl+S**). Confirm the action.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

BIONUMERICS will automatically create the reference system and calibration curve for each of the fingerprint types. Since this allows the calculation of metrics information, a metrics scale now becomes available in the upper part of the *Fingerprints* panel.

Normalization is achieved by assigning bands in the reference channel to external reference positions.

9. To normalize a complete run at once, select ***Normalization*** > ***Auto assign reference positions (all lanes)...*** (, **Ctrl+A**), leave all settings unaltered and press <***OK***>.

10. When the assignment of the marker bands to reference positions is made, the data can be shown in normalized mode with ***Normalization*** > ***Show normalized view*** (, **Shift+N**) (see Figure 10).

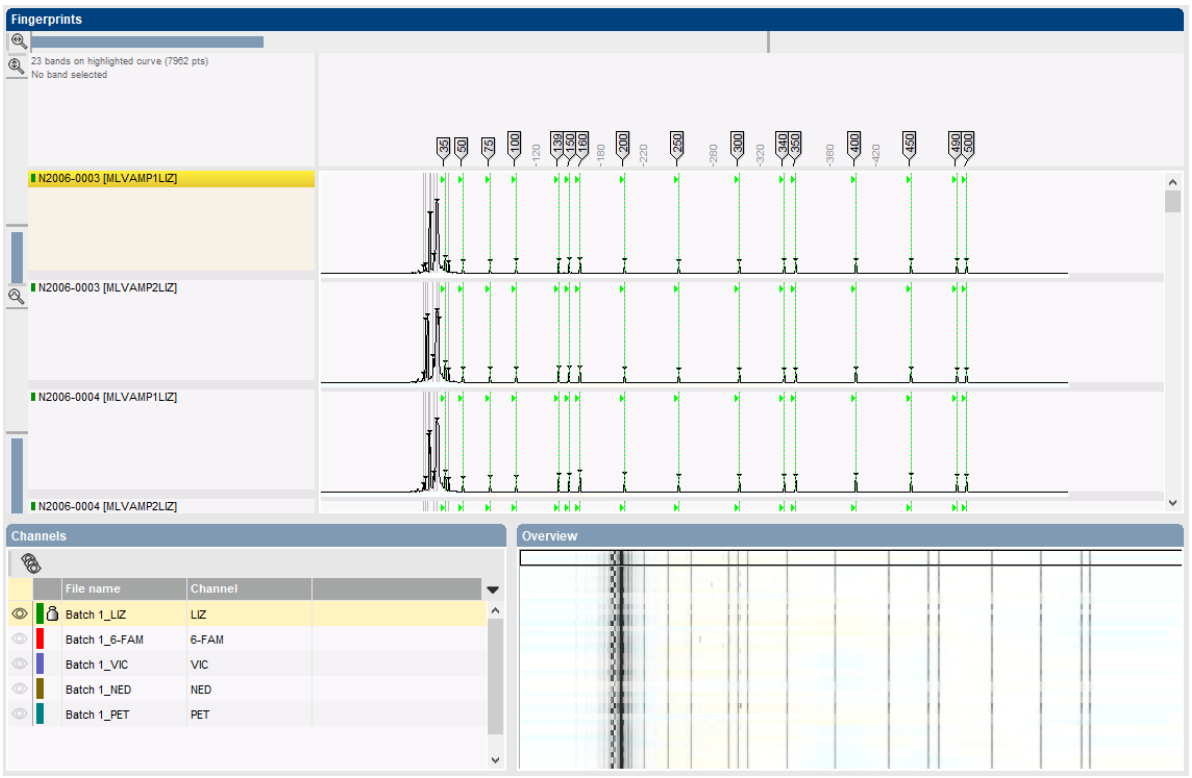

Figure 10: Normalized view - reference dye.

11. Click on the icon left of the data channels in the *Channels* panel.

12. Click on the icon left of the LIZ reference channel.

The data channels are now shown and the reference channel is hidden from the view.

13. Select ***Bands*** > ***Search data bands...*** (, **Ctrl+Shift+F**) to call the *Search data bands* dialog box.

14. Check ***Remove doublets***, ***Remove shadow bands***, ***Filter by fragment length*** and specify a minimum length of 50. Press <***OK***>.

15. Save the changes and close the *Fingerprint curve processing* window. Confirm the new defaults.

The reference system and calibration curve are automatically created for each of the fingerprint types. We can check this from the *Main* window:

16. Double-click on the base fingerprint type in the *Experiments* panel to open the *Fingerprint type* window.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

Figure 11: Normalized view - data dyes.

17. In the *Fingerprint type* window, call ***Settings*** > ***Edit reference system***, or double-click in the *R01 panel* to call the *Fingerprint Reference system* window.

A calibration curve for the reference system **R01** of the base fingerprint type is displayed.

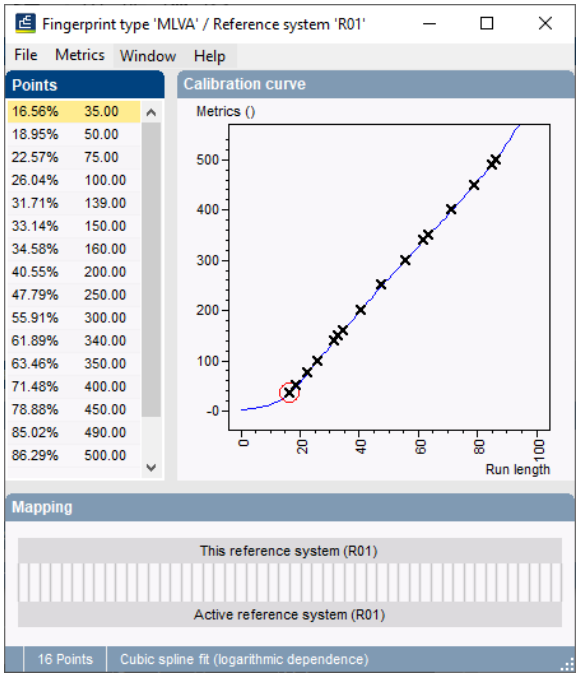

Figure 12: The *Fingerprint Reference system* window, with a calibration curve displayed.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.

18. Close the *Fingerprint Reference system* window and the *Fingerprint type* window.

5 Conclusion

In this tutorial you have seen how to import and process sequencer curve files in BIONUMERICS. You can now start applying comparison functions such as band matching, clustering, etc. on the data. More information about these functions can found in the analysis tutorials on our website.

© 1998-2020 Applied Maths NV. All rights reserved. Not for Diagnostic Use.